Isabelle Audras, Jean-Gabriel Ganascia
LIP6

Titre : Outils d'analyse textuelle d'aide à l'acquisition du français langue étrangère écrit.

1 Cadre théorique

Cette communication croise les domaines de la didactique des langues étrangères, de la stylistique et des sciences cognitives. Les recherches sur l'acquisition de l'écrit en langue étrangère chez l'adulte lettré dans sa langue maternelle sont récentes. Cependant elles ont bénéficié des résultats des recherches en pyschologie cognitive concernant l'acquisition de l'écrit en langue maternelle [Fayol 1996] et en neuro-imagerie concernant l'acquisition de l'écrit chez l'adulte non scolarisé [Morais & Kolinsky 2001]. Prise de recul par rapport à la langue, aide à la mémorisation : les vertus cognitives du passage à l'écrit ne sont plus à démontrer [Mangenot 1998].

De plus, les résultats des recherches en linguistique textuelle croisent également les intérêts de la didactique de l'écrit. En effet les notions de cohérence textuelle et de pragmatique, propre à la linguistique textuelle, facilitent l'acquisition chez l'apprenant d'une compétence textuelle qui lui rend disponible des outils d'articulation pour construire son discours. Autrement dit, « les structures textuelles formelles [...] guident le scripteur dans la construction d'un texte et le lecteur dans sa compréhension » [Scardamalia & Bereiter, 1986].

Tuffs [Tuffs, 1993] renchérit et montre que travailler sur des genres textuels différents facilite l'acquisition de l'écrit en langue étrangère.

2 Problématique

De façon générale, l'écrit en classe de langue est associé à une consigne qui prévoit l'intention de communication, même à l'extérieur d'un genre. En effet, le cadre narratif choisi, par le genre ou la consigne, définit un objectif de communication précis. Celui-ci appelle des objectifs fonctionnels dont l'expression morphosyntaxique et lexicale est vue en classe. Ce contenu linguistique, découvert à l'intérieur d'une situation de communication, est automatisé lors de réemplois, et ceci est d'autant plus vrai si celui-ci se trouve dans un contexte similaire. D'autre part, l'analyse des besoins communicatifs du cadre narratif par l'apprenant aide celui-ci à s'adapter face à une nouvelle situation de communication dans laquelle il doit réagir [Tagliante, 1994].

Par ailleurs, l'apprentissage du FLE est sanctionné par une certification appelée DELF (Diplôme d'Etudes en Langue Française) alignée sur le Cadre Européen Commun de Référence[1] dans l'apprentissage des langues. Les épreuves écrites A1, A2 et B1 du DELF, qui correspondent à trois niveaux d'apprentissage différents, ont pour cadre narratif, respectivement : la carte postale, la lettre amicale, la lettre de motivation.

Dans ces deux types de production écrite (en classe et les certifications) en français langue étrangère, le niveau de l'apprenant est validé par rapport à sa capacité à exprimer un message à travers un modèle appris et reconnu et non simplement par rapport à ses compétences grammaticales.

---

[1] www.ciep.fr

VISIT…

LANZAROTE
Caliente.COM

Ce type d'apprentissage suscite un intérêt scientifique autour de la notion de style produit par l'apprenant. Nous définissons style ici comme l’organisation horizontale des mots. A partir d'un besoin de communication, le style en langue étrangère se définit par l'ensemble des expressions linguistiques choisies parmi celles connues, pour y répondre. Travailler le style en classe de langue contribue à l’enrichissement de la grammaire et du vocabulaire en apportant les nuances nécessaires à une expression juste et variée.
Or, tout apprenant d'une langue étrangère commet des erreurs lors de son apprentissage, qui sont constructives et permettent une progression par leur compréhension et leur correction. De plus, les erreurs de style jouent un rôle d'indicateur du niveau de langue atteint pour l’enseignant [Pery-Woodley 1993], qui sur la base de ces erreurs (types d'erreurs, fréquences de l'erreur) établissent un programme de remédiation. C’est pourquoi nous nous proposons ici de repérer, grâce aux techniques actuelles du traitement automatique des langues, les erreurs usuelles dans le passage à l’écrit d’une population d’apprenants, à travers les deux types de production désignés ci-dessus, ce qui permettra de mettre l’accent, au cours de l’enseignement, sur la correction de ces erreurs.

3 Méthodologie et présentation des outils informatiques

Ce repérage des erreurs peut se faire de deux façons, soit dans l’absolu, par détection des fautes syntaxiques, soit relativement aux usages, par une étude dans un cadre narratif précis des tournures propres à une catégorie d’apprenants, et qui se trouvent absentes ou peu usitées chez les locuteurs natifs. C’est cette seconde approche que nous adopterons ici, sachant que le rôle des enseignants de langue n’est pas d’apprendre une langue abstraite parfaite, mais de transmettre les usages d’une langue.
Enfin, dans un souci d’accompagnement à la production d’un apprenant, nous nous posons la question de la genèse et l’évolution de son style, à travers l’analyse de différentes versions d’une même production.

Plus exactement, le travail présenté ici recourt à l’emploi de deux outils d’analyse textuelle. Le premier est un outil d’analyse stylistique qui dégage les caractéristiques d’une classe d’apprenants, selon leur niveau, et les distingue des locuteurs natifs. Nous avons utilisé dans nos expériences le Littératron mis au point par Jean-Gabriel Ganascia [Ganascia 2001], associé à un analyseur syntaxique, l’analyseur de Vergne en l’occurence [Vergne 2001]. Le second est un outil de comparaison de versions de texte, Medite mis au point par Jean-Gabriel Ganascia [Ganascia, Fenoglio & Lebrave 2004] ; utilisé pour l’analyse de deux versions d’une même production, il aide à comprendre comment l’apprenant gère la construction et la correction de son écrit.

3.1. Présentation du Littératron

Associé à un analyseur syntaxique, le Littératron détecte des motifs syntaxiques récurrents, présents dans un texte et absents d’un autre. Un motif syntaxique est une unité syntaxique cohérente (groupe sujet, groupe verbal, groupe complément). Les analyses linguistiques proposées renseignent l’utilisateur du Littératron (enseignant ou apprenant) sur la présence ou l'absence de tel motif syntaxique. De plus, lorsqu’elles sont confrontées avec une grille de points morpho-syntaxiques attendue pour un certain niveau d’apprentissage - comme c’est le cas pour chaque épreuve écrite de DELF - elles peuvent contribuer à évaluer le niveau de l’apprenant. Nous procédons ici par des analyses comparatives de productions d’apprenants et de francophones natifs.
Deux outils informatiques sont nécessaires pour extraire les motifs syntaxiques caractéristiques des différentes populations. Un motif syntaxique se définit comme une association d’unités linguistiques cohérentes. Voici un exemple de motifs extraits des

analyseurs ayant la structure syntaxique [préposition + pronom personnel réfléchi + verbe à l'infinitif] : "de vous adresser", "afin de vous donner", "de m'investir", "de vous donner". Ces quatre motifs ont été extraits ensemble d'un même groupe de scripteurs de lettres de motivation.

Le premier outil informatique requis est un analyseur morphosyntaxique du français qui construit des arbres syntaxiques à partir de productions écrites. Nous avons eu recourt à l'analyseur linéaire avec dictionnaire partiel Vergne qui a été élaboré par Jacques Vergne de l'Université de Caen, en 1998 (Vergne, 2001). Le deuxième est l'analyseur stylistique Littératron, mis au point au LIP6 par Jean-Gabriel Ganascia (Ganascia, 2001) qui dégage les motifs syntaxiques récurrents présents dans ces arbres.

Plus exactement, à chaque mot ou groupe de mots l'analyseur de Vergne associe une étiquette ; un arbre stratifié est donc une partition d'étiquettes dont les classes dépendent de la profondeur du nœud dans l'arbre d'analyse. Etant donnée une structure d'ASO, le Littératron calcule une mesure de similarité entre plusieurs ASO, fondée sur la notion de distance d'édition, et génère un graphe de similarité enregistrant les sous-arbres les plus proches de l'ASO en entrée.

C'est ce graphe de similarité qui sert ensuite d'entrée à l'algorithme de classification du Littératron, appelé 'centre-étoiles', qui construit des classes de motifs similaires et leur attribue un nom significatif. En effet, l'algorithme centre-étoile évalue d'abord l'ensemble des étoiles centrées sur les différents nœuds puis il prend, pour chacune, la somme des valeurs de similarité des nœuds de chaque étoile au centre. Une fois calculée la valeur de chaque étoile, l'algorithme 'centre-étoiles' prend celle qui à la plus forte évaluation. On marque ensuite, les nœuds qui appartiennent à cette première étoile, avant d'appliquer récursivement le même algorithme sur les nœuds non marqués, jusqu'à épuisement des nœuds non marqués.

En résumé, toute étoile est un sous-graphe du graphe de similarité centré sur un nœud. Pour chaque classe ainsi construite, l'algorithme choisit les motifs les plus similaires au centre de l'étoile, pour illustrer la signification de l'étoile. Il indique aussi le texte source couvert par chacun des motifs.

Voici l'exemple d'un centre d'étoile, illustré par la figure 1 :
[PREP ["de"]] + [GN [ART ["la"]] + [NOM ["forêt"]]] (texte : "de la forêt"), auquel sont associés les 5 motifs syntaxiques suivants :
- [PREP ["à"]] + [GN [ART ["l"]] + [NOM ["auberge"]]] (texte : "à l\'auberge ") ;
- [PREP ["d\"]] + [GN [ART ["un"]] + [NOM ["hiver"]]] (texte : "d\'un hiver") ;
- [PREP ["dans"]] + [GN [ART ["le"]] + [NOM ["monde"]]] (texte : "dans le monde") ;
- [PREP ["avec"]] + [GN [ART ["les"]] + [NOM ["chiens"]]] (texte : "avec les chiens") ;
- [PREP ["depuis"]] + [GN [ADJ ["quelques"]] + [NOM ["jours"]]] (texte : "depuis quelques jours").

Ceci signifie que la mesure de similarité entre le premier motif ('de la forêt') et l'un des arbres dérivés des arbres syntaxiques de chacun de ces cinq groupes nominaux est supérieure à un certain seuil. Ces cinq motifs font partie de la même étoile dont le centre est de la forme : [PREP ["de"]] + [GN [ART ["la"]] + [NOM ["forêt"]]].

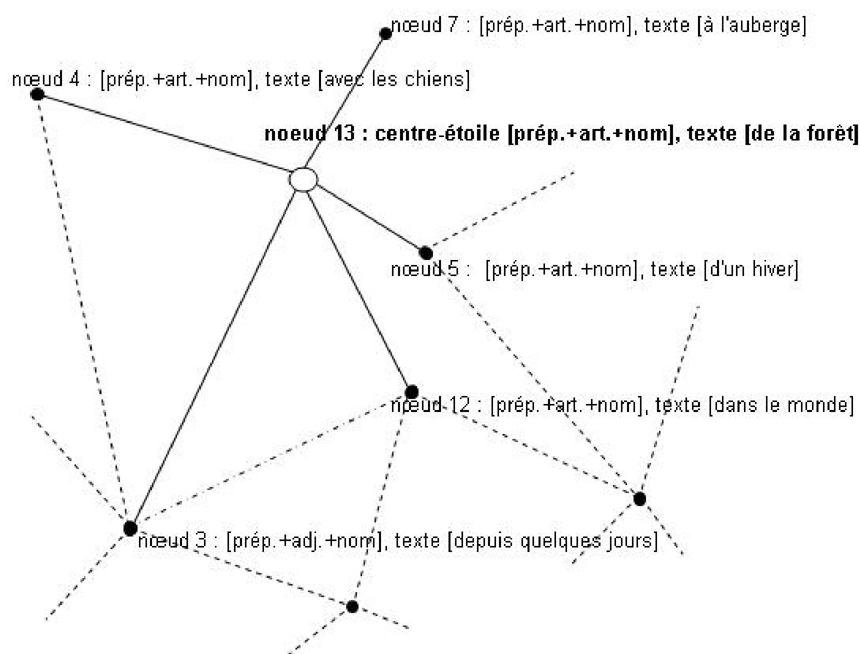


Figure 1 – Graphe du centre-étoile présenté en exemple.

Outre la construction d'étoiles et l'extraction de motifs, le Littératron procède à un second type d'opérations qui consiste à comparer les étoiles issues de plusieurs textes afin de repérer les étoiles présentes dans l'un et absentes dans l'autre. Ceci permet de discriminer, parmi les motifs présents dans une production, ceux qui le distinguent d'autres productions. C'est à partir de ce type de discrimination que l'on construira les tournures caractéristiques de populations d'apprenants.

Par exemple, voici l'une des sorties du Littératron analysant trois groupes de scripteurs distincts à partir de lettres de motivation (LM app : lettres de motivation d'apprenants de langues maternelles diverses, LM appa : lettre de motivation d'apprenants arabophones, LM frcph : lettres de motivation de francophones).

3.2. Présentation de Medite

Médite (Machine pour l'Etude Diachronique et Interprétative du Travail de l'Ecrivain) est un aligneur textuel qui recense systématiquement toutes les transformations qui font passer d'un état textuel à un autre, en mimant les opérations que les généticiens textuels exécutent manuellement. Nous avons choisi de comparer deux types de version de texte : 1. le brouillon de l'apprenant et sa version finale ; 2. une première version rendue à l'enseignant et la version corrigée par l'apprenant suite aux remarques de l'enseignant. Dans ces deux types de situation, cet outil révèle les invariants et les suppressions, insertions et/ou déplacements d'expressions d'une version à une autre, qui sont autant de traces laissées par l'apprenant sur sa démarche d'apprentissage. De plus, l'analyse automatique des motifs syntaxiques les plus fréquemment déplacés, remplacés et/ou supprimés, outil Tree Tagger couplé à Medite, renseigne l'apprenant et/ou l'enseignant sur la qualité stylistique de la production ainsi que sur les points de morpho-syntaxe dont l'emploi est encore hésitant.

4 Présentation du corpus

Des études empiriques conduites autour de trois populations d'apprenants, l'une auprès d'apprenants adultes de langues maternelles diverses en milieu endolingue, l'autre auprès d'apprenants adultes et adolescents arabophones en milieu exolingue et la troisième auprès d'apprenants adultes créolophones en milieu bilingue valident les approches proposées.

Dans le corpus recueilli, nous dénombrons :
- dans chaque population d'apprenants : des productions au cadre narratif fort et relatifs aux certifications DELF telles que la carte postale (A1), la lettre amicale (A2), la lettre de motivation (B1) ainsi qu'une production 'libre' de description d'une même image en couleurs. Parmi ces productions, une partie est faite en classe et l'autre est extraite de copies de DELF.
- chez les apprenants adultes arabophones : des productions argumentatives (examens de cours universitaires)
- chez les apprenants adolescents arabophones : des lettres amicales tirées de copies du DELF scolaire.
- un même nombre de productions répondant aux mêmes consignes, réalisés par des francophones natifs

5 Expérimentations

5.1. Expérimentations avec le Littératron

L'idée de cette recherche est de comparer des productions écrites en classe de F.L.E. de différents niveaux avec des productions de francophones répondant aux mêmes consignes. Les scripteurs francophones sont des natifs français de niveau d'étude au moins équivalent à bac+4. Les apprenants sont de même langue maternelle. Une autre étude, qui pourrait se révéler intéressante, travaillerait avec des francophones d'un niveau d'études moins élevé, voire des débutants scripteurs adultes (Morais J. & Kolinsky R., 2001). L'idée de cette perspective est de montrer que le critère du niveau d'éducation n'est pas négligé dans cette approche.
L'hypothèse est de déceler une spécificité de la syntaxe française au sein d'une classe homogène.

5.1.2. Population arabophone

Présentation des productions écrites

L'ensemble des productions analysées correspond aux examens d'histoire et de civilisation du 1^er^ semestre 2004 d'étudiants en 3^ème^ année du département de français d'une université arabophone. Les étudiants de l'université sont tous de langue maternelle arabe, l'anglais est leur première langue étrangère, le français leur deuxième.
Les productions de chaque type sont au nombre de dix.

Résultats et commentaires

Trois mêmes motifs syntaxiques ressortent systématiquement des productions arabophones. Ces motifs recouvrent ¼ du texte analysé. Il s'agit de deux motifs nominaux et d'un verbal. Les deux motifs nominaux sont de construction : DE + adjectif + nom, comme dans les exemples : 'de choses magnifiques' et 'd'autres villages' :

Impression motif N°184 :

N fp	
det	de
C fp	choses
A fp	magnifiques

Expression couverte : [J’ai vu] de choses magnifiques

Impression motif N°183 :

N ms	
det	d'
A ms	autre
C ms	village

Expression couverte : [Il y avait] d' autre village

Figure 2 – Sortie du Littératron

Ce motif syntaxique révèle une utilisation massive de groupes nominaux compléments du verbe de forme adjectif + nom commençant par DE, au détriment d’autres articles et d’autres prépositions. L’étudiant, à défaut de connaître la bonne rection d’un verbe, directe ou indirecte, ou le bon emploi de l’article partitif sur le défini ou l’indéfini, va utiliser systématiquement la préposition DE pour introduire ses compléments d’objet.
L’apprenant ne semble pas maîtriser une bonne utilisation des prépositions et des articles.
Cette expérience révèle pour une classe homogène d’apprenants arabophones, une spécificité de la syntaxe française dont l’acquisition nécessite un accompagnement particulier.

5.1.3. Population créolophone

Une expérimentation en cours, sur des productions de même consigne, auprès d’étudiants de l’Ecole Normale de Port-au-Prince (Haïti) semble converger vers la même conclusion, avec une flagrante récurrence de motifs extraits de la forme : de + adjectif + nom.

5.2 Expérimentations avec Medite

Présentation des productions

Il s’agit de deux types de production :
- d’une part, deux versions d’une même production : une avant et une après les remarques correctives de l’enseignant (expérience A). Il s’agit d’une critique de film.
- d’autre part, des lettres de motivation de francophones natifs et d’apprenants (expérience B)

Méthodologie, résultats et commentaires de l’expérience A.

Nous soumettons pour comparaison dans le logiciel 2 versions d’une même production, une avant et une après les remarques correctives de l’enseignant.
Les résultats, que nous allons décrire ici, sont présentés figure 3. Sur la figure, à gauche se trouve la 1ère version du texte, à droite la 2ème. Medite utilise 3 couleurs et 1 typologie spécifique pour chacune des 4 opérations qu’il effectue :
- la zone soulignée correspond à ce qui a été déplacé d’un texte à l’autre,
- la zone rouge à ce qui a été supprimé,
- la verte à ce qui a été inséré
- la bleue à ce qui a été remplacé.

En passant d’un texte à l’autre, nous remarquons que :

- nous ne trouvons dans les 2 versions aucun caractère souligné : rien n'a été déplacé d'un texte à l'autre.
- nous trouvons des caractères insérés (vert) uniquement dans le 2ème version du texte et ces caractères sont des marques de correction de lexique ou de grammaire (marque de genre ou de nombre, de terminaisons de verbes conjugués, de rajouts de préposition à forme complexe etc.). Par exemple : dans la version initiale, nous trouvons « grand dépression » et « grande dépression » dans le 2ème version, avec le 'e' de 'grande' surligné en vert.
- nous trouvons des caractères supprimés (rouge) uniquement dans la version initiale du texte. Ils correspondent à des passages mal formulés que l'apprenant préfère supprimer dans la 2ème version, plutôt que d'en rectifier l'expression. Par exemple : dans la version initiale, nous trouvons « [film] aussi très amusant pour les autres a cause d'un bon histoire » ; cette idée a disparu dans la 2ème version.
- nous trouvons des zones de caractères remplacés dans la 1ère et la 2ème version du texte. Ces zones correspondent également à des marques de correction de grammaire ou de lexique. Par exemple : dans la version initiale, nous trouvons « avec les chevals », où les caractères 'ls' de 'chevals' sont surlignés en bleu ; dans la 2ème version, nous trouvons « avec les chevaux », où les caractères 'ux' de 'chevaux' sont surlignés en bleu.

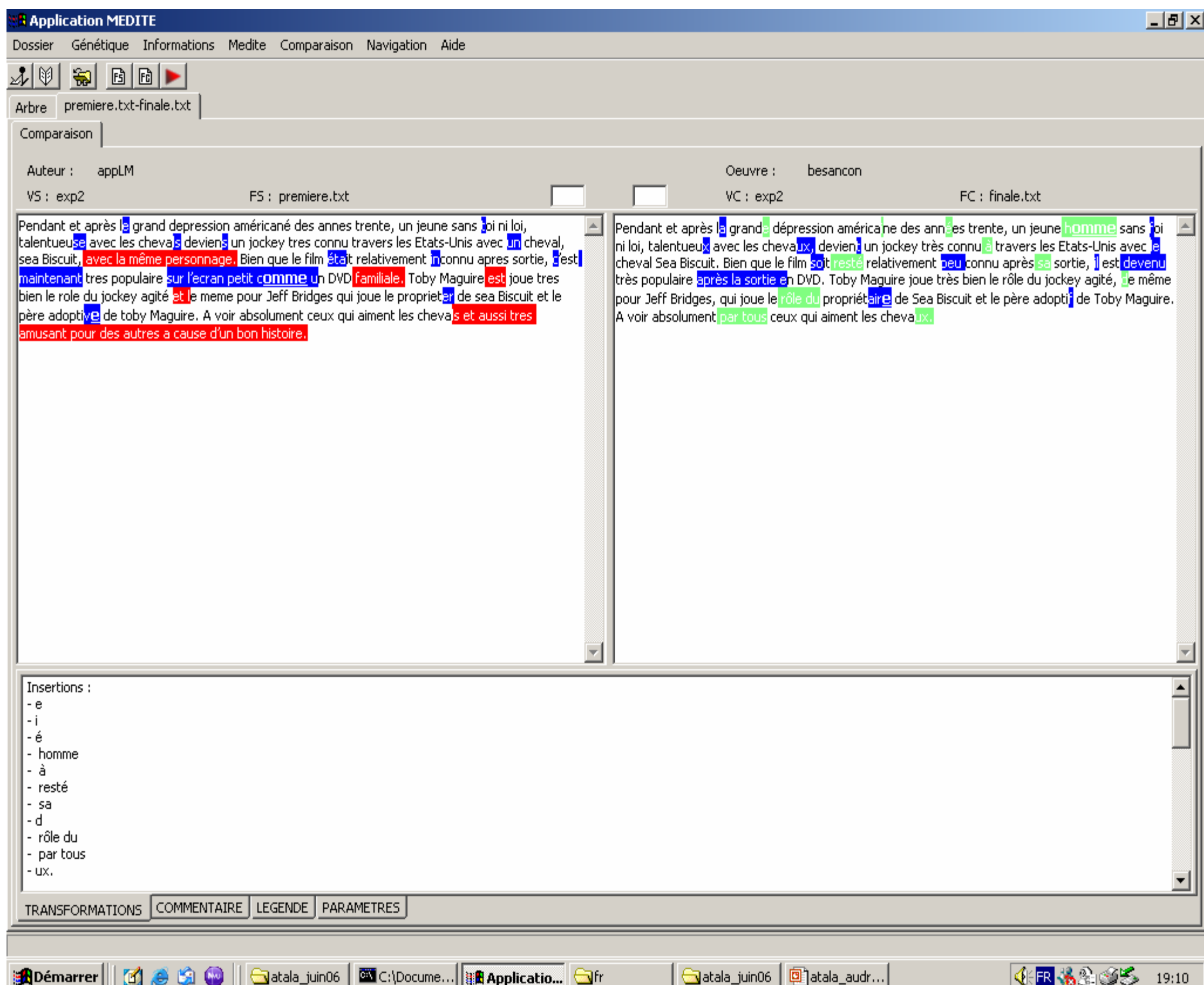


Figure 3 – Copie d'écran de l'expérience A

Medite apparaît bien ici comme un outil neutre, permettant une visualisation du type de bouleversements apportés d'une version à l'autre d'un texte. Ici il s'agit essentiellement de

corrections basiques sur des marques de genre, de nombre ou de conjugaison. L'apprenant a laissé de côté les réécritures de phrases ou d'expressions maladroites.

Méthodologie, résultats et commentaires de l'expérience B.

Nous soumettons pour comparaison dans le logiciel une lettre de motivation d'apprenants (de niveau B1 selon le CECR) et une lettre de motivation d'un francophone natif ; les deux lettres de motivation répondent à la même offre d'emploi.

L'idée de cette comparaison peut paraître saugrenue car nous dévions l'outil de ce pour quoi il a été réalisé, en l'occurrence comparer deux versions d'un même texte, donc d'un même auteur. Par cette comparaison, nous ne cherchons pas à faire ressortir les caractéristiques d'une 'bonne' version (ici celle du scripteur natif) par rapport à une 'mauvaise' (ici celle de l'apprenant). Nous partons de l'hypothèse que dans ce type d'écrit très formaté, autant sur le plan discursif (cadre narratif) que linguistique (lexique et morpho-syntaxe), les expressions d'une lettre à l'autre vont peu varier. Nous allons interroger les variances extraites, non pas sur la plan de la grammaticalité avec l'œil de l'enseignant qui corrige les fautes et rappelle la norme, mais plutôt sur celui de la littérarité avec l'œil de l'esthète qui cherche fréquences, rythme, choix de l'expression dans le jeu aléatoire de sa lecture. Cet objectif peut sembler un peu hardi ; disons que nous cherchons à lever le voile sur ce qui pourrait être appelé un style en langue étrangère, avec toutes les restrictions que ce terme contient[2]. Nous verrons que l'exemple proposé ne permet qu'en partie de révéler la notion de style en langue étrangère ; d'autres expériences sont en cours sur ce sujet.

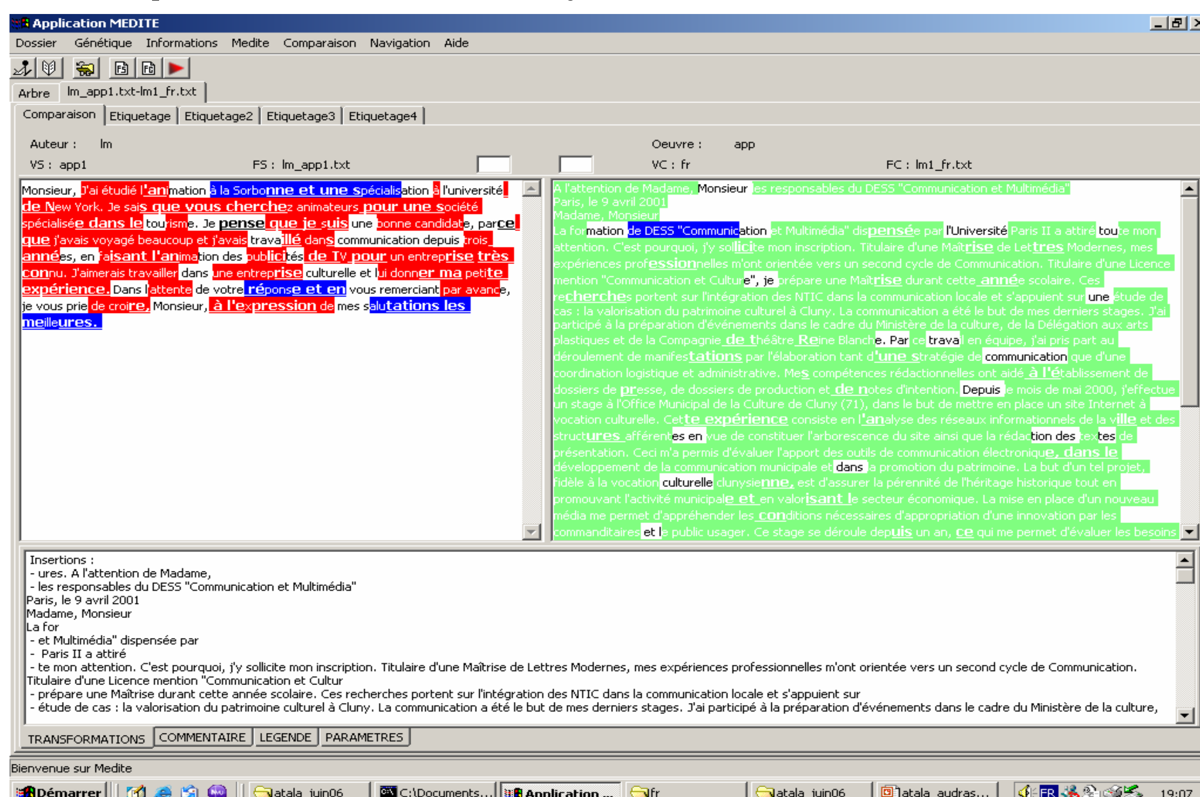


Figure 4 – Copie d'écran de l'expérience B

[2] Nous cherchons ici à 'positiver' les informations extraites en changeant de point de vue ; celles-ci ne sont plus vues en tant qu'expressions erronées ou maladroites d'un apprenant-scripteur mais plutôt en tant qu'expressions singulières d'un apprenant-auteur. [Derrida 67]

Les résultats, que nous allons décrire, sont présentés figure 4. Medite utilise les mêmes couleurs pour spécifier les 4 opérations qu'il effectue que pour l'expérience A. Sur la figure, à gauche se trouve la lettre de motivation de l'apprenant, à droite celle du francophone.
Nous remarquons :
- à droite, dans la lettre de motivation du scripteur natif, une très grande quantité de texte surlignée en vert (insertion) ; à gauche, dans la lettre de motivation de l'apprenant, une grande quantité de texte surligné en rouge (suppression)
- une petite quantité de textes remplacés (zone bleue)
Apparemment, les deux lettres sont donc très différentes dans l'expression. En les lisant l'une derrière l'autre, nous observons qu'elles ont pourtant bien le même objectif et sont bien situées dans le même contexte. L'expression est différente : le choix des mots, leur agencement dans le texte, etc.

Pour appuyer notre courte étude, nous nous servons de l'outil d'analyse syntaxique Tree Tagger, également disponible dans Medite. Celui-ci, par des mesures statistiques, nous renseigne sur le pourcentage de chaque type de mots utilisés (ou étiquettes classées selon la nature syntaxique du mot : adjectif, verbe, substantif, préposition etc.). Nous trouvons :
- dans la lettre du scripteur natif, une plus grande proportion de noms, d'adjectifs, de prépositions et de déterminants ;
- dans la lettre du scripteur apprenant, une plus grande proportion de verbes, d'adverbes, de conjonctions ;
- dans les deux lettres, une même proportion de signes de ponctuation
A vue d'œil, le texte de l'apprenant est plus court que celui du francophone natif.
Ces résultats quantitatifs nous permettent de conclure que :
- les phrases de l'apprenant sont plus courtes, davantage centrées autour du noyau verbal
- les phrases du francophone natif sont plus longues, davantage tournées autour de noyaux nominaux (noms et ses extensions)
Nous définissons ici les premières caractéristiques de ces deux scipteurs-auteurs. Nous croisons ces caractéristiques avec d'autres informations :
- c'est l'une des premières lettres de motivation en français produites pour l'apprenant en question ; il a par ailleurs suivi une formation essentiellement tournée sur la compréhension et la production orale, il a donc peu l'habitude de l'écrit ; il est en contexte de préparation à l'écrit du Delf niveau B1
- le francophone natif est un étudiant de master, très habitué à l'écrit et qui a déjà postulé à des emplois et envoyé des lettres de motivation similaires à celle présente ici.
Nous ne souhaitons pas questionner ici les différentes hypothèses qui amènent l'auteur-apprenant à utiliser des phrases courtes mais complexes (présence de conjonctions) : est-ce parce qu'il n'est pas sûr de ses capacités à l'écrit ? Peut-être manque-t-il d'habitude pour ce genre d'exercice ? Est-il stressé dans la perspective de l'examen ? Notre propos est uniquement d'ouvrir sur la possibilité d'autre lecture des informations extraites par Medite.
Une étude longitudinale, sur un corpus plus élargi, nous renseignerait sur d'autres caractéristiques (fréquence des mots utilisés, rythme etc.) et complèterait de révéler les deux styles en présence.

L'expérience B semble indiquer que Medite est un outil d'extraction d'informations propres à nourrir le questionnement autour de la qualité stylistique d'un écrit en langue étrangère, en dehors d'une problématique de repérage des erreurs et d'aide à la correction. Ainsi les informations extraites automatiquement sont autant de traces qui aident à la compréhension du processus d'apprentissage, dans ce qu'il a d'individuel et de personnel, comme le style

d'un écrivain l'est pour la compréhension du processus de création, également dans ce qu'il a d'individuel et de personnel.

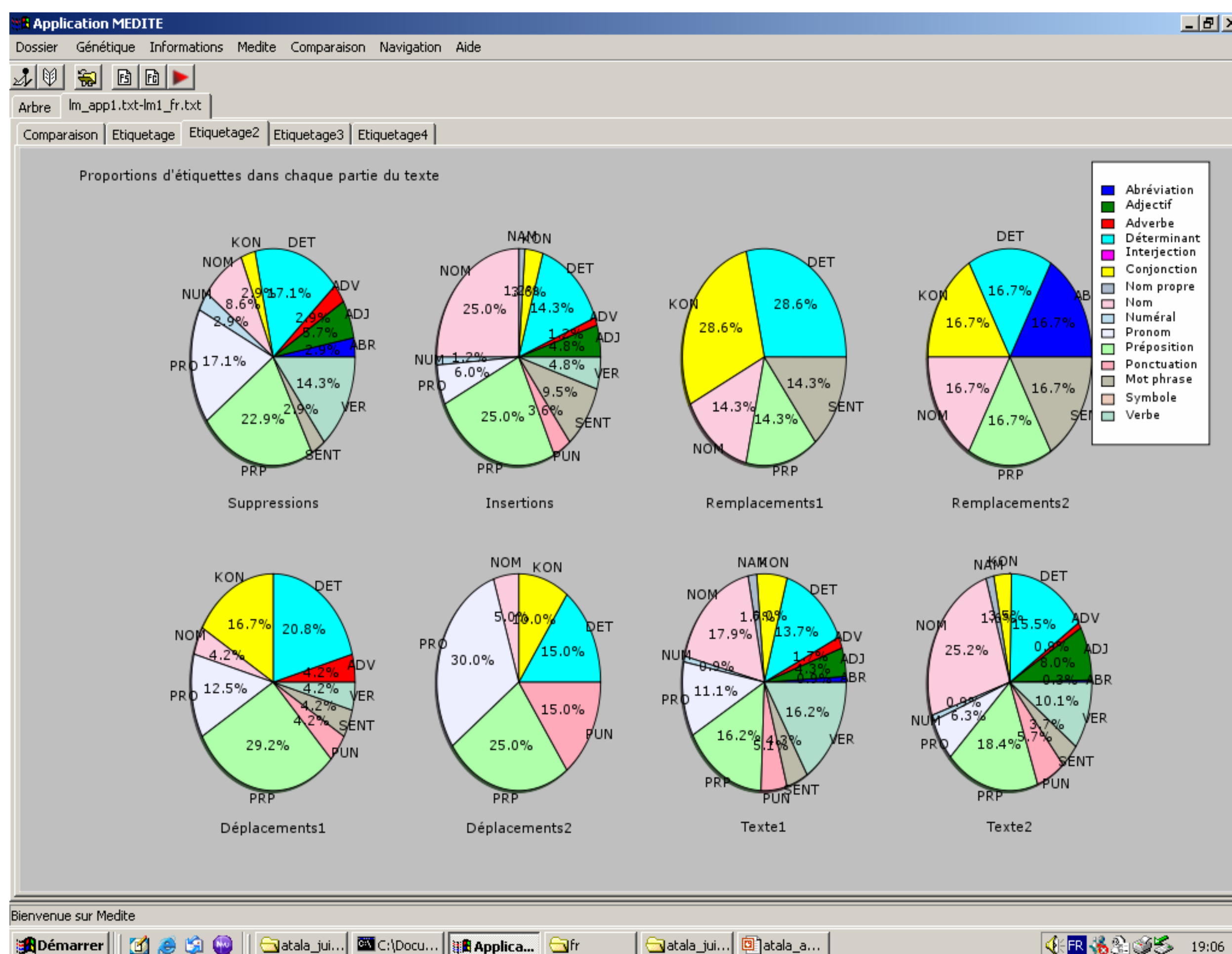


Figure 5 – Copie d'écran des résultats d'étiquetage obtenus avec Tree Tagger dans l'expérience B.

6. Conclusion

Utilisé en sciences du langage dans le domaine de l'acquisition en langue étrangère du français écrit, le Littératron et Medite se révèlent être des outils pertinents d'aide à la rédaction. En effet, le Littératron est en mesure d'effectuer un diagnostic linguistique de l'apprenant sur des productions au cadre narratif contraignant. Ce diagnostic sert à l'accompagnement à l'acquisition de l'écrit.
Medite permet de visualiser la production écrite au moment de sa révision. Accompagnée d'un tuteur, cette visualisation aide l'apprenant à prendre conscience de ses erreurs et la verbalisation qui accompagne cette conscientisation aide à l'apprentissage. D'autre part, l'analyse des informations extraites automatiquement par Medite aide à la réflexion autour de la notion de style en langue étrangère.

A terme, ce travail doit faire l'objet de deux types de développements complémentaires, sur les plans technique et expérimental.
D'un côté, nous nous sommes limités ici à une décomposition en syntagmes, et à une étude de la structure de la phrase relativement à cette décomposition. Cela restreint assez fortement le type de motifs détectés. Nous allons faire appel à une décomposition plus riche qui prendra en

compte la structure propositionnelle. L'algorithme d'extraction de motifs est identique, mais l'analyse syntaxique diffère ; surtout, l'arbre résultant de cette analyse doit être considérablement enrichi.
D'un autre côté, les résultats obtenus auprès d'étudiants arabophones et créolophone nous encourage à poursuivre plus loin l'étude des différences spécifiques auprès d'apprenants venant de différentes régions du monde, et de langues maternelles diverses, sur des productions de DELF.

Références

Barthes R. (1972). *Le degré zéro de l'écriture*. Paris : Seuil.

Derrida J. (1967) *De la grammatologie*. Paris : Les Editions de Minuit

Dupoux, E. (ed.) (2001). *Language, brain and cognitive development : Essays in Honor of Jacques Mehler*, Cambridge, Mass. : MIT Pr. Morais J. & Kolinsky R. (2001). "The literate mind and the universal human mind". In Dupoux. pp. 463-480.

Fayol, M. (1996). La production du langage écrit. In J. David & S. Plane (Eds.), *L'apprentissage de l'écriture de l'école au collège*. (9-36). Paris : Presses Universitaires de France.

Ganascia J.-G., Fenoglio I., Lebrave J.-L. (2004), EDITE MEDITE: un logiciel de comparaison de versions, *actes des 7ième journées d'Analyse statistique de Données Textuelles (JADT 2004)*, pp. 468-478

Ganascia, J.-G. (2001). "Extraction automatique de motifs syntaxiques". *In actes du colloque Traitement Automatique du Langage Naturel 2001* (*TALN 2001).* Consulté en juillet 2005. http://www.li.univ-tours.fr/taln-recital-2001/index1.html

Mangenot, F. (1998). "Outils textuels pour l'apprentissage de l'écriture en L1 et en L2"**.** In *Pratiques discursives et acquisition des langues étrangères*, Souchon, M. (dir.). *Actes du Xe colloque international "Acquisition d'une langue étrangère : perspectives et recherches"*. Besançon : Université de Franche-Comté. pp 515-525.

Pery-Woodley, M.-P. (1993). *Les écrits dans l'apprentissage*. *Clés pour analyser les productions des apprenants*. Paris: Hachette FLE.

Scardamalia, M. & Bereiter, C. (1986). "Research on written composition". In *Handbook of research on teaching*, Wittrock, M.C. (dir.). New York : McMillan. pp 778-803.

Tagliante, C. (1994). *La classe de langue.* Paris : CLE International.

Tuffs, R., (1993), "A genre approach to writing in the second language classroom : the use of direct mail letters". *Revue belge de philology et d'histoire*, vol. 71, n°3, pp. 691-721.

Vergne, J. (1999). Etude et modélisation de la syntaxe des langues à l'aide de l'ordinateur. Analyse syntaxique non combinatoire Synthèse et résultats. Habilitation à Diriger des Recherches, Université de Caen. http://users.info.unicaen.fr/~jvergne/#HDR.

Vergne, J. (2001). Analyse syntaxique automatique de langues : du combinatoire au calculatoire. Actes de *Traitement Automatique du Langage Naturel 2001* (*TALN 2001). http://users.info.unicaen.fr/~jvergne/Taln2001FR_JV.pdf*